Algorithmic aspects of multicanonical simulations

Bernd A. Berg[a]

[a]Department of Physics and Supercomputer Computations Research Institute,
Florida State University, Tallahassee, FL 32306, USA

Monte Carlo (MC) simulations of many systems, in particular those with conflicting constraints, can be considerably speeded up by using multicanonical or related methods. Some of these approaches sample with a-priori unknown weight factors. After introducing the concept, I shall focus on two aspects: (i) Opinions about the optimal choice of weight factors. (ii) Methods to get weight factor estimates, with emphasize on a multicanonical recursion.

1. INTRODUCTION

One of the questions which ought to be addressed before performing large scale computer simulations is "What are suitable weight factors for the problem at hand?". It has been expert wisdom for quite a while [1], and became widely recognized in recent years [2-8], that MC simulations with a-priori unknown weight factors are feasible and deserve to be considered.

The design of suitable weights requires to understand or guess relevant physical features of the system(s) under (numerical) investigation. This is already exemplified by the work of Metropolis et al.[9]. To sample the important configurations of a canonical ensemble at temperature $T = \beta^{-1}$ efficiently, the Boltzmann weight is recommended

$$w(a) = e^{-\beta E_a}\,, \tag{1}$$

where E_a is the energy of configuration a. Namely, it is straightforward to show that this weight minimizes the sampling error for the canonical (normalized) energy density $P(E)$:

$$\sum (\triangle P(E))^2 = \text{minimum}\,. \tag{2}$$

However, even when studying canonical statistical mechanics the Boltzmann weight is not a satisfactory recipe for all cases. It can be false that the weight factor (1) samples the important configurations. A well-studied case is the sampling of configurations with interfaces [2,15]. Quite generally, instead of (2) the sampling error for the actually calculated observable needs to be optimized.

A second reason, called "dynamic" in [8], relies on the fact that the error estimate (2) holds for independent configurations. To sample with a weight like (1), one has to rely on a Markov chain. This introduces an autocorrelation time which depends on the weights chosen. In this sense the argument leading from (1) to (2) is "static" [8]. It can happen that statically most efficient weights are rendered inefficient by the dynamics, whereas statically worse weights can yield a favorable dynamics. For instance, in systems with conflicting constraints encouraging experience has been made with weights [3,4] and expanded ensembles [10–13] which allow to "refresh" the system in the disordered phase. It seems worthwhile to study this concept also for less difficult situations, like second order phase transitions.

In the next section I consider the question of optimal weights. The third (and major) section focuses on how to get working estimates of the weights in the first place. Brief conclusions are given in section 4.

2. WHICH WEIGHTS?

First, I consider the static aspects. For the relative sampling error of the spectral density $n(E)$ we get

$$\sum (\triangle n(E)/n(E))^2 = \text{minimum} \tag{3}$$

by using

$$w(a) = 1/n(E_a)\,. \tag{4}$$

The result is an uniform energy probability density. For a large number of applications this has proven to be a robust choice, for instance when dealing with first order phase transitions [2,15]. Instead of the energy, weights in other thermodynamic variables have also been used successfully, for example $w(a) = f(m_a)$ [16], where m_a is the magnetization of configuration a, or $w(a) = f(b_a)$ [5], where b the bound variable for cluster updating.

In some applications one is tempted to introduce weight factors in two variables, like $w(a) = f(E_a, m_a)$, which would allow to re-construct canonical expectation values over both, a macroscopic temperature and magnetic field range. Only for exceptional cases, when one is interested in very small volumes, this seems to be practical. In most cases the interest lies in a finite size scaling (FSS) analysis of the infinite volume limit $V \to \infty$. CPU time slows down worse than V^2 for the discussed methods. One cannot afford to replace V by V^2, i.e. a performance worse than V^4. In addition, a RAM increase with V^2 is unpleasant too. The art seems to be to find *one* combination of variables, such that suitable weighting with it leads to significant improvements of the numerical performance.

With respect to groundstate searches it is claimed in ref.[6] that the weight

$$w(a) = 1/\sum_{E'=E_{\min}}^{E_a} n(E') \tag{5}$$

gives the best worst case performance in terms of ergodicity and pertinence. The authors define pertinence by the static argument that VN_s independent samples with the weight (5) are better than N_s independent rival samples. Ergodicity relates to the dynamics of the Markov chain. The latter is system dependent and there is presently little understanding of the processes driving its slowing down. Hence, it seems to me that their statement about ergodicity is unwarranted.

The idea of expanded ensembles [10,11], particularly attractive in their version of parallel tempering [14,13], tries to overcome dynamical slowing down by introducing moves into additional dimensions. Whereas new weights may alter or by-pass barriers, expanded (canonical) ensembles can only by-pass them, because Boltzmann weights are used for the normal updates. A, yet untested, possibility would be to apply the parallel tempering idea to unconventional ensembles. This may allow to combine the advantages of both approaches.

3. RECURSIVE WEIGHT ESTIMATES

Weight factors like (4) or (5) are a-priori unknown. This introduces some additional complications, unknown to canonical simulations. Before the simulation can be performed, a working estimate of the weight factor has to be known, *i.e.* an approximation for the practical purpose of performing the MC simulation. Deviations by factors around ten may well be acceptable, because the accuracy of the simulation is hardly affected by this. The errors of the simulation remains in any case statistical, because the actually used weight factors are exactly known. On the other hand, deviations from the desired weights by factors of, say, order 10^{10} (well the order of improvements) are certainly no longer acceptable. Such large discrepancies would entirely obstruct the basic ideas.

In some cases, for instance first order phase transitions [16], FSS allows reasonable working estimates on the basis of already simulated smaller systems. Also constraint simulations have successfully been used [17,18]. For other situations, like spin glasses [3], this does not work. In the remainder of this section I will first explain the difficulties of the most straightforward recursion. Subsequently, I give an efficient solution, which essentially is a simplified and improved presentation of an result of ref.[19].

Desired is a working estimate of the weight factor (4). Others can be treated similarly. It is advisable to start any recursion in the disordered phase of a system, where the system moves most freely under MC updates. Therefore, let us assume for the zeroth simulation.

$$w^0(a) = 1 \text{ for all } a\,. \tag{6}$$

The most obvious recursion goes now as follows: Simulation n, $(n = 0, 1, 2, ...)$ is carried out with the estimate $w^n(a)$ and yields the histogram $H^n(E)$. Estimate $n + 1$ for the weight factors is then given by

$$w^{n+1}(a) = \frac{w^n(a)}{H^n(E_a)}\,. \tag{7}$$

This recursion has a number of difficulties

(i) What to do with histogram entries $H^n(E) = 0$ or small?

(ii) Each $H^n(E)$ calculation starts with zero statistics. Assume, someone has given us the exact result and we use it for $w^0(a)$: The next estimate $w^1(a)$ will be worse. A noisy left-over of $w^0(a)$.

(iii) Initial weights $w(a) = 1$ correspond to temperature infinity and are bad in the limit $E_a \to E_{\min}$, approached towards zero temperature.

To avoid (i), one may replace

$$H(E) \to \hat{H}(E) = \max\left[h_0, H(E)\right], \tag{8}$$

where h_0 is a number ≤ 1. The recursion derived in the following overcomes (ii).

Let us first discuss the relationship [20,19] of the weight factors with microcanonical temperature $\beta(E)$, because it turn out to be advantageous to derive the final recursion in this quantity. It holds

$$w^{n+1}(a) = e^{-S(E_a)} = e^{-\beta(E_a)\,E_a + \alpha(E_a)} \tag{9}$$

where $S(E)$ is the microcanonical entropy and, by definition,

$$\beta(E) = \frac{\partial S(E)}{\partial E}\,.$$

This determines the fugacity function $\alpha(E)$ up to an (irrelevant) additive constant. Consider, for example, the case of a discrete minimal energy ϵ. We may choose

$$\beta(E) = \left[S(E+\epsilon) - S(E)\right]/\epsilon \tag{10}$$

and the identity $S(E) = \beta(E)\,E - \alpha(E)$ implies $S(E) - S(E-\epsilon) = \beta(E)E - \beta(E-\epsilon)(E-\epsilon) - \alpha(E) + \alpha(E-\epsilon)$. Inserting $\epsilon\,\beta(E-\epsilon) = S(E) - S(E-\epsilon)$ yields

$$\alpha(E-\epsilon) = \alpha(E) + \left[\beta(E-\epsilon) - \beta(E)\right]\,E \tag{11}$$

and $\alpha(E)$ is fixed by defining $\alpha(E_{\max}) = 0$. In summary, once $\beta(E)$ is given, $\alpha(E)$ follows for free. The starting condition (6) becomes (other $\beta^0(E)$ choices are of course possible)

$$\beta^0(E) = \alpha^0(E) = 0\,. \tag{12}$$

With $\hat{H}$ defined by (8) we translate now eqn.(7) into an equation for $\beta(E)$. Subscripts $_0$ are used to indicate that those quantities are not yet the final estimators from the n^{th} simulation. Let

$$w_0^{n+1}(E) = e^{-S_0^{n+1}(E)} = c\,\frac{w^n(E)}{\hat{H}^n(E)}\,,$$

where the (otherwise irrelevant) constant c is introduced to ensure that $S_0^{n+1}(E)$ can be an estimator of the microcanonical entropy. It follows $S_0^{n+1}(E) = -\ln c + S^n(E) + \ln \hat{H}^n(E)$. Inserting this relation into (10) gives

$$\begin{aligned}\beta_0^{n+1}(E) = \beta^n(E) + \\ \left[\ln \hat{H}^n(E+\epsilon) - \ln \hat{H}^n(E)\right]/\epsilon\end{aligned} \tag{13}$$

As estimator of the variance follows

$$\sigma^2[\beta_0^{n+1}(E)] = \frac{c'}{H^n(E+\epsilon)} + \frac{c'}{H^n(E)}\,,$$

where c' is an unknown constant and we have re-introduced the original histograms to emphasize (and use) that the variance is infinite when there is zero statistics. The statistical weight for $\beta_0^{n+1}(E)$ is inversely proportional to its variance. Choosing a convenient constant, we get

$$g_0^n(E) = \frac{H^n(E+\epsilon)\;H^n(E)}{H^n(E+\epsilon) + H^n(E)}\,. \tag{14}$$

Note that $g_0^n(E) = 0$ for $H^n(E+\epsilon) = 0$ or $H^n(E) = 0$. The n^{th} simulation was carried out using $\beta^n(E)$. It is now straightforward to combine $\beta_0^{n+1}(E)$ and $\beta^n(E)$ according to their respective statistical weights into the desired estimator:

$$\beta^{n+1}(E) = \hat{g}^n(E)\,\beta^n(E) + \hat{g}_0^n(E)\,\beta_0^{n+1}(E)\,, \tag{15}$$

where the normalized weights

$$\hat{g}_0^n(E) = \frac{g_0^n(E)}{g^n(E) + \hat{g}_0^n(E)} \text{ and } \hat{g}^n(E) = 1 - \hat{g}_0^n(E)$$

are determined by the recursion

$$g^{n+1}(E) = g^n(E) + g_0^n(E),\ g^0(E) = 0\,. \tag{16}$$

We can eliminate $\beta_0^{n+1}(E)$ from eqn.(15) by inserting its definition (13) and get

$$\begin{aligned}\beta^{n+1}(E) = \beta^n(E) + \hat{g}_0^n(E) \times \\ [\ln \hat{H}^n(E+\epsilon) - \ln \hat{H}^n(E)]/\epsilon\end{aligned} \tag{17}$$

Notice that the sole purpose of using $\hat{H}$, instead of H has become to avoid $\ln(0)$. The weight $\hat{g}_0^n(E)$ of such contributions is zero. The major advantage of (17), compared with (7), is that it keeps already assembled statistics. This solves (ii) and eases (iii). Frequent iterations are allowed, implying increased stability and decreased CPU time consumption. In addition one may implement a suitable $\beta(E)$ guess for $E \to E_{\min}$.

Finally, eqn.(17) can be converted into a recursion for ratios of weight factor neighbors. We define

$$R^n(E) = e^{\epsilon\,\beta^n(E)} = \frac{w^n(E)}{w^n(E+\epsilon)} \tag{18}$$

and get the recursion

$$R^{n+1}(E) = R^n(E) \left[\frac{\hat{H}^n(E+\epsilon)}{\hat{H}^n(E)}\right]^{\hat{g}_0^n(E)}\,. \tag{19}$$

4. CONCLUSIONS

Sampling of broad energy and other unconventional distributions is technically feasible. Statical slowing down can be overcome, as is well established for first order phase transitions. Dynamical slowing down is more difficult, but further algorithmic improvements (find the relevant combination of variables!) appear to be possible.

Acknowledgement: I would like to thank Wolfhard Janke for useful discussions.

REFERENCES

1. G.M. Torrie and J.P. Valleau, J. Comp. Phys. 22 (1977) 187.
2. B.A. Berg and T. Neuhaus, Phys. Rev. Lett. 69 (1992) 9.
3. B.A. Berg and T. Celik, Phys. Rev. Lett. 69 (1992) 2292.
4. U. Hansmann and Y. Okamoto, J. Comput. Chem. 14 (1993) 1333.
5. W. Janke and S. Kappler, Phys. Rev. Lett. 74 (1995) 212.
6. B. Hesselbo and R. Stinchcombe, Phys. Rev. Lett. 74 (1995) 2151.
7. G.R. Smith and A.D. Bruce, Phys. Rev. E53 (1996) 6530.
8. B.A. Berg, in *Proceedings of the International Conference on Multiscale Phenomena* (Bielefeld, Sept. 30 – Oct. 4, 1996), edited by F. Karsch, B. Monien and H. Satz, World Scientific, 1997, pp.137–146.
9. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys. 21 (1953) 1087.
10. A.P. Lyubartsev et al., J. Chem. Phys. 96 (1992) 1776.
11. E. Marinari and G. Parisi, Europhys Lett. 19 (1992) 451.
12. W. Kerler and P. Rehberg, Phys. Rev. E50 (1994) 4220.
13. K. Hukusima and K. Nemoto, J. Phys. Soc. Japan 65 (1996) 1604.
14. M.C. Tesi et al., J. Stat. Phys. 82 (1996) 155.
15. W. Janke, in *Computer Simulation Studies in Condensed Matter Physics VII*, edited by D.P. Landau, K.K. Mon and H.-B. Schüttler, Proceedings in Physics 78, Springer, 1994, pp.29–44.
16. B.A. Berg, U. Hansmann and T. Neuhaus, Z. Phys. 90 (1993) 229.
17. A. Billoire, T. Neuhaus and B.A. Berg, Nucl. Phys. B396 (1993) 779.
18. F. Csikor et al., Phys. Lett. B357 (1995) 156.
19. B.A. Berg, J. Stat. Phys. 82 (1996) 323.
20. B.A. Berg and T. Neuhaus, Phys. Lett. B267 (1991) 249.